The uncertainty principle of visual color evaluation

HAGEN, Eddy, independent researcher
Contact and info: eddy.hagen@insights4print.ceo, https://www.linkedin.com/in/eddyhagen/

Abstract: It is a common practice, both in research settings and in daily life: visual color evaluations. In research experiments on e.g. color discrimination, participants are asked whether they see a difference between two samples. In the printing industry e.g., it frequently happens to have a 'press check' where customers are asked to evaluate the color quality of a print when comparing it with a proof. But is that a reliable practice? Based on experiments and observations in real life, this paper challenges the premise that one can be objective in the visual evaluation of color if one knows he/she is evaluating color. As the experiment shows: if one knows the task is to evaluate color, psychology might overrule the objective assessment of color and people start seeing things that are not there: color differences between identical samples. Ergo: the uncertainty principle of visual color evaluation.

Keywords: color, color appearance, color evaluation, brand colors, printing, packaging, print quality

Introduction

Both in general printing and packaging printing quality control has historically largely depended on visual inspection. Although measurement devices and ISO-standards have gained adoption worldwide and provide better and more consistent quality control, it still happens that a customer demands a 'press check' to be sure that the print job has the 'right' color, the 'right' quality. Often a proof is used as a reference for this visual evaluation. And often this is considered a 'contract proof': based on the color appearance of both proof and print, a print buyer might decide whether to accept or reject a print job. Which means that this visual evaluation comes with significant responsibility.

But how reliable is visual color evaluation? And not only in research and press checks, but also on supermarket shelves? Some believe that consumers won't buy a product if the color deviates, even a tiny bit. This believe translates into stringent quality requirements, allowing only small color deviations, much lower than ISO-standards describe.

Research questions

This research attempts to answer the following four questions:

1. Which color deviations are visible?
2. Which color deviations are considered 'disturbing'?
3. Which color deviations would influence buying behavior?
4. Is there a difference in results between flat samples (resembling press checks) and folded samples (resembling supermarket shelves)? And is there a difference between 'bleeding' and 'non-bleeding' images on folded samples?

Experiment setup

To investigate the questions above, the following experimental setup was used:

1: a small package was designed, on one side it had a bleeding image (colors up to the border of the box), the other side had a non-bleeding image. The packages resemble the small cereal boxes one finds at the breakfast buffet in hotels.

2: in total, eight variations were made, the reference had a 100% M (magenta) and 100% Y (yellow) 'brand color', the others had lower percentages of M and/or Y.

	Ref	1	2	3	4	5	6	7
C	0	0	0	0	0	0	0	0
M	100	95	90	100	100	95	90	100
Y	100	95	90	95	90	100	100	97
K	0	0	0	0	0	0	0	0

3: the boxes were printed in offset, on the same substrate as breakfast cereal boxes, at the end of an actual order of this kind of

boxes. The same finishing was applied: coating, die-cutting. A part of the test set was also folded and glued, the other part was delivered as flat (unfolded) samples.

4: the differences between the variations and the 'reference' (100M, 100Y) were measured with a calibrated spectrophotometer (X-Rite SpectroEye).

5: a test procedure was designed. Participants would evaluate the samples in a portable light booth (standardized conditions) and answer the following questions for each variation compared to the reference: do you see a difference, do you find that difference disturbing, would a difference like this influence your buying behavior? This was done both for the flat samples and the folded samples, both bleeding and non-bleeding.

What makes this research unique is that also an identical copy was included in the test set: at one time in the three parts (flat, folded/bleeding, folded/non-bleeding), the participant would compare two identical samples. Which proved to be a valuable test.

6: evaluation of the results

The design of the experiment setup was done by the author of this paper, the execution was done by a bachelor student as part of his bachelor proof (Jens Adriaensen, AP College, Antwerp – Belgium, 2015). The evaluation of the test results for this paper was again done by the author.

Test set
Below is a picture showing the design of the test set and the setup. During the test, always two samples were shown: the reference plus one variation.

(image by Jens Adriaensen)

The different variations had the following measured deviations when compared to the reference, as already stated, one of them was identical:

- dE00 = 0 (dE76 = 0)
- dE00 = 0,5 (dE76 = 1,0)
- dE00 = 0,9 (dE76 = 1,7)
- dE00 = 1,2 (dE76 = 2,2)
- dE00 = 1,7 (dE76 = 3,7)
- dE00 = 2,1 (dE76 = 3,4)
- dE00 = 3,2 (dE76 = 5,3)
- dE00 = 3,4 (dE76 = 6,3)

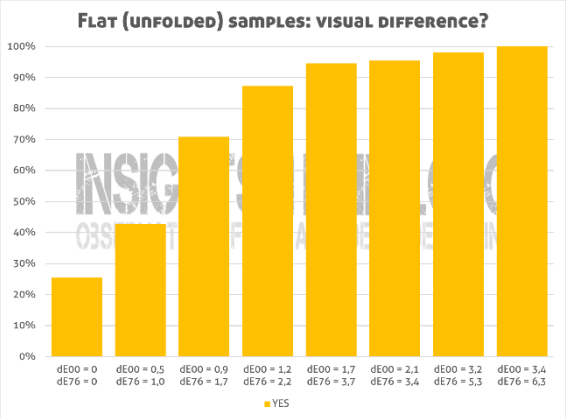
It should be mentioned that the measured colors and deviations differ from the theoretical values based on the design: a lot of parameters influence the reproduction of a color in print.

Participants
This is a breakdown of the people that participated in the test.

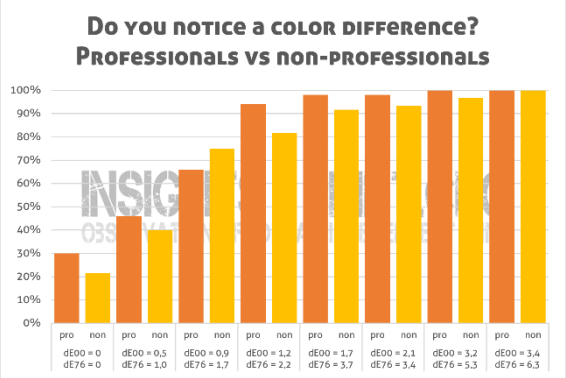
- 55% female; 45% male
- 51% some involvement with print production; 49% no relation with print production
- Age: 34% 18-25y; 25% 26-35y; 27% 36-50y; 14% 51-65y; no one was younger than 18 or older than 65.

Experiment results: flat samples
The first research question should have been rather straightforward, but it wasn't. Already after a dozen tests, something strange showed up: someone 'saw' a deviation

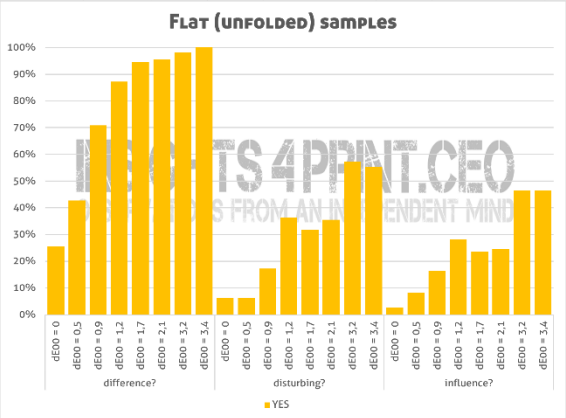
between the two identical samples. And this was not an outlier: at the end of the experiment, it showed that 25% of the participants claimed to see a difference between the two identical samples.



When comparing the people who had some involvement with print ('print professionals') with the others, the number of people that claimed to see a difference was higher with the print professionals.

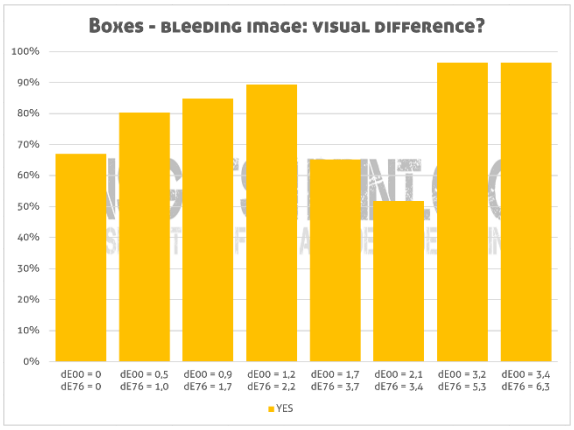


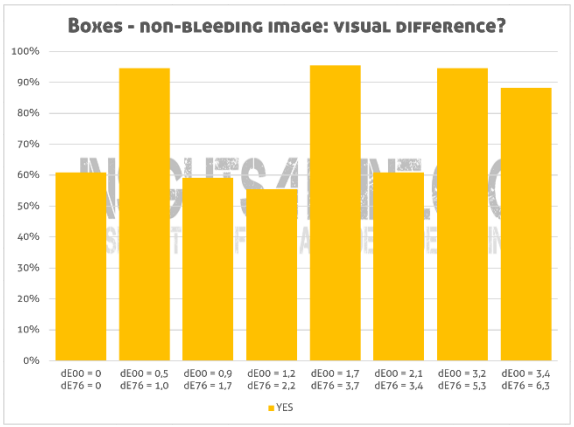
As expected, seeing a difference doesn't mean that it is considered disturbing, that it would influence buying behavior. The graph below shows the relation between the different questions.



Test results: folded samples

The surprising result of the test with the flat samples was even too a higher degree present in the folded samples: 60% (non-bleeding image) to 67% (bleeding image) claimed to see a color difference between the two identical samples.

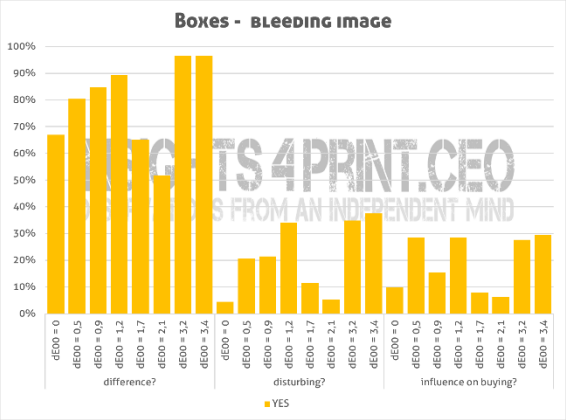


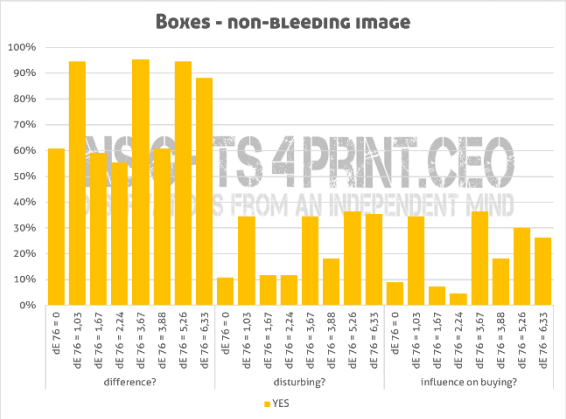


And next to that high percentage of people seeing color differences between identical samples, the correlation between the number of people noticing color differences and the measured results seemed to be gone. The results of the folded boxes are inconsistent with the measured color deviation, contrary to the unfolded samples where the numbers were increasing as expected.

Next to that: even between the 'bleeding' and 'non-bleeding' boxes, there is no relation.

With these results, one could question whether the results of the other two questions are still relevant: the participants didn't see the deviations as expected. But for completeness, below are the results.





Discussion

Flat samples

If people see a difference where there is none, that is a liability for press checks: one could reject a perfect valid job. Also for color research, this is a liability: it undermines the validity of the experiments. So, what could be the reason for these findings?

The most plausible one is the question asked: subjects were literally asked: "Do you notice a difference, on a scale from 0 (no difference) to 3 (big difference)?", which implied that they had to focus on finding a difference, eventually fooling many of them. This could also explain why more print professionals claimed to see a difference between the identical copies: they are professionals, they know color, it's in their job description to be critical about color. Therefore, they look at color in a very different way than non-professionals, they want to see a difference.

This has a considerable consequence for both press checks and color research: a part of the participants performing a visual color evaluation can't be trusted. Especially in color research, an 'objectivity' check should always be included. And all people who do not pass this test, by claiming to see a difference between two identical copies, should be eliminated from the test results.

For this research, eliminating the unreliable results was an after-thought. At the moment this thought first occured, the original data set wasn't available anymore.

Folded samples

With a majority of participants claiming to see a color difference between the identical samples, and seeing no correlation between the positive evaluations and growing color deviation, other factors than the psychological effect of the phrasing of the question must have played a role.

Since the unfolded samples did show the expected relation, the reason can't be found in the kind of deviation (e.g. changes in hue), as someone who looked at the data suggested: if that were true, all three parts of the test should have shown the same pattern, which is not the case, they all differ.

The most probable cause is related to the samples and their placement. The boxes were folded, which means that the front side probably wasn't 100% flat but slightly curved, which might lead to differences in color appearance and gloss. Also, the two samples might not have been in precisely the same plane, they might have been under a slightly different angle, resulting in somewhat different lighting and therefor slightly different color appearance, gloss.

Please note: this reflects real-life, this is also the case on supermarket shelves. Which does make this finding very relevant: it's impossible to accurately compare colors in this kind of setting. And therefore: if consumers aren't able to correctly evaluate colors in a supermarket, could they be able to distinguish the 'right' brand color from a 'wrong' one and leave the 'wrong' one on the shelves? Especially if the deviation is small?

Short real-life experiment
The effect of the formulation of research questions, even revealing the fact that research is related to color, is shown in the following short experiment.

These two boxes of cereal bars were next to each other in the supermarket. For someone involved with color, it's immediately clear that the yellow is different.

The following discussion shows a conversation with a female, late forties, university degree, who was shown the two boxes.

Q: Do you notice anything?
A: Euh… you bought two packages of granola bars… Are you hungry?
Q: No, that's not it. Do you notice anything about those two packages?
A: Euh…
Q: Do you notice a difference between those two packages?
A: That one has a dent in it, it is damaged.
Q: No, that's not it. Do you notice anything about the "Golden Syrup" text?
A: Euh… not really…
Q: Do you notice anything about the color of the background behind the "Golden Syrup" text?
A: It's a nice yellow…
Q: No, that's not it. Do you notice a color difference between that background on the two packages?
A: Euh… well, yes, now that you point my attention to it, there seems to be a slight color difference.

Although this is only a short experiment with only one participant, it does point to an important fact: for people who are not professionally involved in color, in print production, they will solely focus on color when pointed explicitly towards it.

Conclusions
The experiment shows that a significant number of people are not objective in color assessment if they know they are evaluation color. This has serious consequences both for color research and common practices like 'press checks': the uncertainty principle of visual color evaluations.

The experiment also shows that color perception, color appearance off non-flat objects and objects that are not exactly in the same plane, should be further investigated. This research suggests that even slight changes in placement or curvature might result in noticeable color deviations, even if the color of the packages is identical.

Further research
Confirmation of the findings of this research would be interesting and recommended.

Next to that, further research on 'real-life' color would be interesting: color discrimination of non-flat objects (e.g. red Coca-cola cans) and color discrimination of colors that are not in the same plane.

In the first case, the shadow where the cans are the closest to each other might influence color discrimination, making it more difficult to notice small color deviations.

The second case is when two identical sample boxes are placed under a slightly different angle, one perpendicular to the viewing direction and the other at varying degrees to the first one, e.g. a series of 5, 10 and 15 degrees. Would this influence color appearance and color discrimination? This could prove or reject the assumption made above. Since boxes in supermarket shelves are not perfectly aligned, this is very relevant for brands that insist on exact color appearance.

PS: the concept of 'the uncertainty principle of visual color evaluation' was first published in August 2017 on the insights4print-blog: https://www.insights4print.ceo/2017/08/uncertainty-principle-visual-color-evaluations/